

# The First Intron in the Human *c-abl* Gene Is at Least 200 Kilobases Long and Is a Target for Translocations in Chronic Myelogenous Leukemia

ANDRÉ BERNARDS,<sup>1\*</sup> CHARLES M. RUBIN,<sup>2</sup> CAROL A. WESTBROOK,<sup>2</sup> MICHAEL PASKIND,<sup>1</sup>  
AND DAVID BALTIMORE<sup>1</sup>

Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142<sup>1</sup>; Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139<sup>1</sup>; and Joint section of Hematology/Oncology, Department of Medicine, The University of Chicago, Chicago, Illinois 60637<sup>2</sup>

Received 13 April 1987/Accepted 11 June 1987

**The *c-abl* protooncogene is unusual in two respects; it has multiple, widely spaced N-terminal coding exons transcribed by different promoters, and it is the target of the translocations that form the Philadelphia chromosome found in cells of chronic myelogenous leukemia patients. To understand the organization of the gene in normal and chronic myelogenous leukemia patient DNA we have mapped *c-abl* by pulsed field gradient gel electrophoresis. We find that one of the alternative 5' exons of the gene lies at least 200 kilobases upstream of the remaining *c-abl* exons, posing formidable transcription and splicing problems. The 5'-most *c-abl* exon includes an unusually long 1,276-base-pair segment that contains 15 ATG codons and multiple short open reading frames, upstream of the *abl* initiator codon. Its peculiar structure suggests that *c-abl* may be decapitated in most chronic myelogenous leukemia patients, and we demonstrate that this is the case in the chronic myelogenous leukemia cell line K562.**

The *c-abl* protooncogene is the cellular homolog of *v-abl*, the oncogenic element of the Abelson murine leukaemia virus (13). That virus causes a unique pre-B-lymphocytic leukemia in mice (1). The *c-abl* gene is a member of the tyrosine kinase family of protooncogenes (14, 31), is widely expressed in the body (30), and is highly conserved from humans to *Drosophila melanogaster* (19). The *c-abl* gene is also the site of the first consistent tumor-associated chromosomal translocation ever found, the Philadelphia chromosome translocation of human chronic myelogenous leukemia (CML) (17, 25).

The *c-abl* gene has a unique organization. This became evident after we cloned four types of murine *c-abl* cDNA (*c-abl* types I to IV [4]). These cDNAs were colinear with each other from the poly(A) tail upstream through most of the *abl* coding region, but diverged at a common position just short of the *abl* N terminus, and the cDNAs predict four *abl* proteins with different N termini (4). Two of the cDNAs represent the two major mouse *c-abl* mRNAs of 5.3 kilobases (kb) (type I) and 6.5 kb (type IV). The two remaining types represent minor mRNAs. The point where the cDNAs diverge corresponds to a splice site in the mouse genome. The mRNAs represented by the cDNAs, therefore, appear to consist of alternative 5' exons added onto a set of common 3' exons (4). Shtivelman and co-workers have recently reported similar results for the human *c-abl* gene (29). These authors cloned cDNA copies of the two human *c-abl* mRNAs and located the type I exon, which they called exon 1a, 19 kb upstream of the common exons. The human type IV sequence (exon 1b) was also cloned, but was not present within 35 kb upstream of exon 1a (29). Because the equivalents of the mouse *c-abl* type II and III mRNAs have not yet been identified in humans, we will follow the exon 1a, 1b nomenclature in this paper. Exons 1a and 1b are each preceded by a transcriptional promoter in both humans and

mice (29; unpublished results). The two major *c-abl* mRNAs in mammals are thus initiated by separate promoters.

The possibility that exon 1b of the human *c-abl* gene lies far upstream of the remaining *c-abl* exons is significant because the region upstream of the common exons is thought to be interrupted by a chromosomal translocation in the vast majority of patients with CML. In this translocation, the common *c-abl* exons and a very variable length of DNA upstream are translocated from their normal location on chromosome 9 (band q34) to a 22q<sup>-</sup> or Philadelphia chromosome (17). In the CML cell line K562, for example, more than 97 kb of DNA upstream of the common exons is cotranslocated (16). By contrast to this variability in the position of the chromosome 9 breakpoint, the Philadelphia chromosome breakpoints are clustered on chromosome 22 (15). This breakpoint cluster region (*bcr*) is part of a gene, which in Philadelphia chromosome-positive CML is transcribed into a hybrid *bcr-abl* mRNA (12, 28). The hybrid mRNA encodes a P210 *bcr-abl* fusion protein, which is thought to play an important role in the pathogenesis of CML (21).

Because of the variability in the position of the translocation breakpoint near the *c-abl* locus, it has remained unclear whether the *c-abl* gene is consistently disrupted in Philadelphia chromosome-positive CML. To shed light on this issue we have mapped approximately 600 kb around the human *c-abl* gene by pulsed field gradient gel electrophoresis. We report that the 5'-most exon of *c-abl* lies at least 200 kb upstream of the common exons and that the gene is disrupted in the CML cell line K562.

## MATERIALS AND METHODS

**Isolation and characterization of exon 1b clones.** Genomic exon 1b clones were isolated by screening a human recombinant DNA library with a 0.95-kb *EcoRI* mouse *c-abl* type IV cDNA fragment under conditions of reduced stringency (50% formamide, 5× SSC (1× SSC is 0.15 M NaCl plus 0.015

\* Corresponding author.

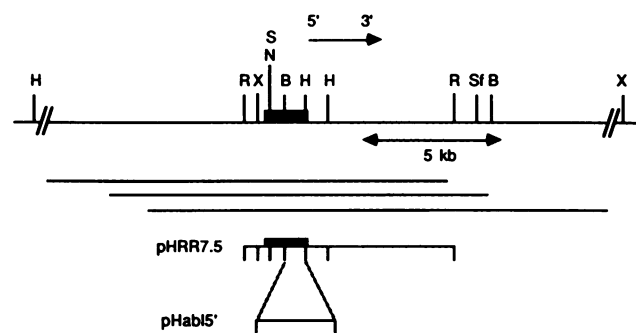


FIG. 1. Restriction map of the *c-abl* exon 1b genomic region. The exon is drawn as a black box; the direction of transcription is indicated. Three phage clones that represent this region of the human genome are drawn beneath the restriction map. From one of these phage a 7.5-kb *EcoRI* fragment was subcloned to yield plasmid pHRR7.5; a subclone of this plasmid (pHab15') contains a 0.85-kb *BamHI-HindIII* fragment that is internal to exon 1b. The insert of this plasmid was used as a human exon 1b probe in the experiments reported here. Abbreviations of restriction enzymes: B, *BamHI*; H, *HindIII*; N, *NotI*; S, *SacII*; Sf, *SfiI*; R, *EcoRI*; X, *XhoI*. The *HindIII* and *XhoI* fragments detected by the exon 1b probe are 15 and 18 kb long, respectively. Not all *BamHI* and *HindIII* sites in the exon 1b region are drawn.

M sodium citrate), 37°C hybridization; 2× SSC, 50°C wash). The recombinant library contained size-selected (>15-kb) *MboI* partially digested fragments of normal human DNA in the phage cloning vector EMBL3 and was provided by Nathaniel Landau. Positive clones were plaque purified and subcloned in plasmid vectors by standard procedures. The DNA sequence of exon 1b was established by the chemical degradation method (24) on both strands. Five base-specific modification reactions were employed to minimize ambiguities.

**Pulsed field mapping of *c-abl*.** High-molecular-weight DNA was prepared by incorporating SupT3 (a human T-cell leukemia line) or RCH-ACV (a human pre-B-lymphoid line) cells into blocks of low-melting-point agarose, followed by in situ lysis and deproteinization. This procedure has been described in detail elsewhere (5). The DNA was digested with restriction enzymes while still incorporated in agarose (5), and the blocks of agarose containing the digested DNA were transferred to the slots of a 1% agarose gel and subjected to pulsed field gradient gel electrophoresis as described by Carle and Olson (7). Briefly, gels were run with two alternating 250-V electrical fields alternating every 30 s for 24 h. The size-separated DNA was transferred to GeneScreen and hybridized as described previously (5). Probes were labeled by the random priming method (11).

## RESULTS AND DISCUSSION

The two major transcripts of the human *c-abl* gene consist of alternative 5' exons (exons 1a and 1b) spliced onto common 3' exons. Exon 1a (the type I exon) lies 19 kb upstream of the first common exon. Exon 1b (type IV) does not map within 35 kb upstream of exon 1a (29). To complete the map of the human *c-abl* gene we isolated exon 1b by screening a human genomic library with a mouse *c-abl* type IV cDNA probe. Three positive phage clones were isolated. A restriction map of the exon 1b region is presented in Fig. 1. A 7.5-kb *EcoRI* fragment that contains exon 1b was subcloned, and the DNA sequence of exon 1b was deter-

mined on both strands (Fig. 2). Our sequence differs from the previously reported partial exon 1b sequence (29) in 21 positions (Fig. 2). Some of these differences may be DNA sequence polymorphisms or cloning artifacts; others may be sequencing errors. Also shown are four mRNA start sites that were mapped by Shtivelman et al. by RNase protection (29; labeled "a" to "d" in Fig. 2). The region 5' of site "d" contains at least four potential binding sites for the transcription factor Sp1 (10) and two CCAAT sequences (underlined in Fig. 2). We do not find a typical TATA sequence in this region. However, the sequence TTAAAAAGG 59 base pairs downstream of the 3' CCAAT box (underlined in Fig. 2) resembles the CATAAAAGG element, which functions as a TATA box in the rabbit beta-globin gene (9). Whether these sequence motifs are part of the upstream *c-abl* promoter is under investigation.

The 45-residue *abl* N-terminal amino acid sequence predicted by exon 1b starts with the ATG codon at position 2027 in the exon 1b sequence (Fig. 2). The human amino acid sequence differs in two positions from the mouse type IV sequence (4). Both amino acid sequence changes (Lys for Arg at position 29 and Glu for Asp at position 30) are conservative substitutions. The human and mouse sequence do not differ in the last codon before the exon 1b splice donor site (at position 2163 in Fig. 2), which encodes a histidine residue in both species. We make note of this because this residue is absent in the previously reported human sequence (29). Our sequence does not show a second splice donor site that could give rise to an alternatively spliced mRNA variant lacking the histidine codon, and we have no explanation for this discrepancy. Interestingly, exon 1b includes 1,276 bp upstream of the *abl* initiator codon. This part of the mRNA contains multiple short open reading frames (Fig. 2). Only the 5'-most of the 15 ATGs upstream of the *abl* start codon conforms reasonably well to Kozak's consensus sequence for eucaryotic initiator codons (PuCCATGG [22]). The remaining 14 upstream ATGs and the *abl* initiator codon show no particularly good homology to this consensus (Fig. 2). Computer analysis of the exon 1b sequence reveals some imperfect hairpin loops and direct repeats. Analysis of the exon 1b sequence in another species may reveal whether these structures are evolutionarily conserved. The exon 1b sequence is not significantly homologous to any sequence in current data bases.

To determine how far upstream of the common *c-abl* exons exon 1b lies, we used the observation that the former and at least 97 kb of the upstream DNA are amplified approximately fourfold in DNA from the CML cell line K562 (8, 16). The amplified chromosome 9;22 breakpoint region in this cell line also includes the immunoglobulin lambda light chain constant region (27) and the 5' part of the *bcr* gene. Upon hybridizing a blot containing *BamHI* digests of Raji and K562 DNA with a probe that represents exon 1a and the first two common exons, both the 3.5-kb fragment containing the common exons and the 15-kb exon 1a fragment were clearly at an increased molarity in K562 DNA compared with Raji DNA and were therefore both amplified in K562 (Fig. 3A). By contrast, a 7.5-kb exon 1b exon fragment was not increased in K562 and therefore was not amplified (Fig. 3A). We conclude that *c-abl* exon 1b lies upstream of the chromosome 9;22 translocation breakpoint in K562 cells. Because more than 97 kb of DNA upstream of the common *c-abl* exons has been cotranslocated in K562 (16), exon 1b must lie further upstream.

To determine the distance between exon 1b and the first common exon of *c-abl*, we used the Carle and Olson

CCTCAGGTGATCCGCCGCTCTCGGCCTCCCAAAGTCTGGGATTACCGCGTGAGCCACCGCGCCGCCGCTAGTTCTAATATTTATGCTTTTGGAGGGTGAAGGGCTTGAAGGCA 120  
 TCAAAAACCTTGGTTTCTAGGAGCGAGCCAAATTTTCTCCCTTAGAGACTGAAGGAGTTATCCGACTCCGCTTCCAGGAAATCCAATCCGCAATTTTCTCCCTCGCTTGGCTCTTCCA 240  
 GGTCTCACTACTCGGCCGGGAGACGCGGTCCGGCTACGCTAGGCCGGCAAGTCTACCAGGTCCAGGAGGGGGGGCCGAGCCCAAGCGCGGGGCCGGCCAGGAGTCTT 360  
 GGTGCAGGCGGACTCGGCCGCTGCCATGGGACCAACAATCCCGCACGCACCGCGGACTGCGCCGAACACGGTTGGCTGGGAGATGCCAACCTACTGCTCCACCGGCCAATGGGACG 480  
 GAGGAAAGAGGCTGACCCGGATTCTCGGAGGCCAATGGCAATACAGAGGAGCGGGTGTCTCTTACCCTCACCTTGCCTCAGAGAAGCCACATTTAAAGGTCGTCCGGGGCCGGG 600  
 CTTTCGACGGCAGTGGGACAGAGCCGGGAGGGGGGGGGTGTCCCGGAAAGCGGCTCTTGGGGCATCCGGGTCCCTCGAGGGGGTCCGGACGTCCGGTGGGGGGCCGAGGGCG 720  
 X "d" M A A A G F G V C A \* T  
 GGGCCTGGCCTCGTTGTGGAGCGGCTCGTAATCCATCATGGCGGCCCGGGGTTTGGTGTCTGTGCCTGAGCAGCGCTGGAGCCGAGCCGG-TTCCCGGGTCTCGGGCTGAGGAGCCC 839  
 X M \*  
 CTCCGCTGTCCACGGCCCTACCGGGGGGGGGGGTGGTCCCTCGCGGAGCTCGGGAGATGTGACTGCCTGAGGGCGGTGGTGGTGTGAGCGTCCGGGGCCGGGGAGGGGGTGTCT 959  
 "c" "b" A "a" X  
 CGGCAGAGACCCCGGGTCTGGGCGAGCTGAGGCGGCCGGGCTCCTCTACACGGGGCCCGCTTCCGCTGTCTGGGCCGAGAGTCTTCGTCCCTTACAGCCCGCCCGGCTTGT 1079  
 A  
 GGACACTGCGGGTGGTCTGTTTCCCCAGCTTGGGACACCCGTTTCTGAGGCGTGAAGAGC-GTCCGCCCGAGTAAGCTGCCGTGCCGCGCCCGACAGCTTCCCTCAGCCCCAA 1198  
 X X X X M  
 GCGGCCCTTAATCCGGATCCGGGCCCAACTTTGGCCACGGAGCCTCCCATCAATCCCTCCCTTGTGTCAAGGGGTCTCCCTTCCCCAAGTGGCTCCCGGAGCCTCTAATGC 1318  
 S \* M S P T A P L V S A Q P K T L M Q R K S L D W L H R P F K K R T \*  
 CCTGACTTCTTCAATGTACCTACGGCCCCCTTAGTCTCAGCTCAGGCAAAATCTTAATGCAAGGAAAGTCTGGATTGGCTCCACAGGCCTTTAAAGCGGACTTAAAGTGC 1438  
 M H S F S S E S R A N S L K S G \*  
 TGGCAATGCATTCTTTTCGTGAGTCAAGGCAAACTCGCTGAAATCTGGGTGACCCGTGTCTTTTCCGAGAGCAAGAGAGAGAGCGAGCGGCCACTAGTTCGGCAGGAAAT 1558  
 M K K L R \* M C E N \* M T S I  
 TGTGGAAGATGAAGAAGCTAAGATAGGGGGTGGTGAATTCACAGGAAAGTCTGGAGGAGTAGCCAAAGACCATCAGCGTTTCTTTATGTGTGAGAATTGAAATGACTAGCATT 1678  
 I D P F Q H P L \* GC M  
 TTGACCCTTTTCAGCATCCCTGTGAATATTTCTGTTAGGTTTCTTCTTGAAGAAATGTTATTAGCCCGTTTAAACAAATCAAGAACTTTGGGTAAACATTGCAATTACAT 1798  
 GA M S F C D \*  
 K L I T A K I I G T P A C K C Q P K K S A S F C Y G R C L S V I D F N C \*  
 GAAATTGATAACCGGAAATAATTTGAATCTCTGCTGCAAGTGTCAACCTAAAAAAGTGTCTCTTTGTTATGGAAGATGCTTTCTGTGATTGACTTCAATTGCTGACTTGTGGA 1918  
 G M E D V F L \*  
 M Q R M \* M P F P S T L A D R S G R S \* CTC M G Q Q  
 GATGCAGCGAATGGAATCCACGATATGCAATTTCCCTCTACGCTCGCTGACCGTTCTGGAAGATCTTGAACCTCTCTGGAAGGGGTACCTATTATTACTTTATGGGCAGCAG 2038  
 P G K V L G D Q R R P S L P A L H F I K G A G K K E S S R H G G P H C N V F V E  
 CCTGGAAGTACTTGGGACCAAGAAGGCCAAGCTTGCCTGCCCTGCATTTTATCAAAGGAGCAGGGAAGAAGGAATCATCGAGGCATGGGGTCCACACTGCAATGTTTTGTGGAA 2158  
 H  
 CATG GTGAGTGTCTTTCAAATTTCTGCTCATGTTTTCTCATGCATTATCTTAGGCCTTCAAGGAACCTTGAACAACAGTACTTGCACAGTTCCTTCAATTCACCTTAATAAAT 2277  
 XXX splice

FIG. 2. DNA sequence of human *c-abl* exon 1b and flanking segments. The first 60 base pairs of this sequence consists of an Alu repeat. Shtivelman and co-workers have reported that exon 1b-containing transcripts start at either of four sites. These sites are labeled "a" to "d." The position of site "d" is approximate only (29). Four potential binding sites for the transcription factor Sp1, two CCAAT boxes, and the TTTAAAAGG sequence upstream of site "d" are underlined. The 15 ATG codons between transcription initiation site "d" at position 751 and the *abl* start codon at position 2027 are also underlined. The short open reading frames that follow these start codons are shown. A partial exon 1b sequence, corresponding to nucleotides 689 through 1233 and 1685 through 2159 in the present sequence, had been previously reported (29). The 21 differences between our sequence and the sequence reported by Shtivelman et al. are indicated (X indicates nucleotides that were absent from the latter sequence).

modification of the pulsed field electrophoresis technique, originally described by Schwartz and Cantor (7, 26). Recognition sites for the restriction enzyme *Sfi*I that are cleavable in human DNA were mapped by standard procedures 25 kb downstream of the first common *c-abl* exon and 6 kb downstream of exon 1b (Fig. 4). A human *c-abl* genomic probe (A28-2-2; Fig. 3B legend) that maps upstream of the *Sfi*I site in the common exon region hybridized to a 225-kb *Sfi*I fragment (Fig. 3B). The exon 1b probe did not hybridize to this fragment but, instead, hybridized to a 75-kb *Sfi*I fragment (data not shown; Fig. 4). Exon 1b must, therefore, lie at least 200 kb upstream of the common exons. *Not*I and *Sac*II sites map just within both exon 1a and 1b (Fig. 4) and are cleavable in human DNA. The exon 1b probe recognized

175-kb *Not*I and 150-kb *Sac*II fragments (Fig. 3B), neither of which was seen with common exon probes (Fig. 4), confirming the presence of a large intron in the *c-abl* gene. To test whether the 175-kb *Not*I fragment seen with the exon 1b probe extends up to the *Not*I site in exon 1a (which would fit with the 225-kb *Sfi*I fragment extending up to the *Sfi*I site downstream of exon 1b), the pulsed field blots were re-probed with probe BB0.9. This 0.9-kb *Bam*HI-*Bgl*II fragment, which maps approximately 10 kb upstream of exon 1a, indeed hybridized to a 175-kb *Not*I fragment (data not shown; Fig. 4). We think, therefore, that exons 1a and 1b of the human *c-abl* gene are 175 kb apart, although if there are multiple 175-kb *Not*I fragments the distance could be longer. In the former case, the intron between exon 1b and the first

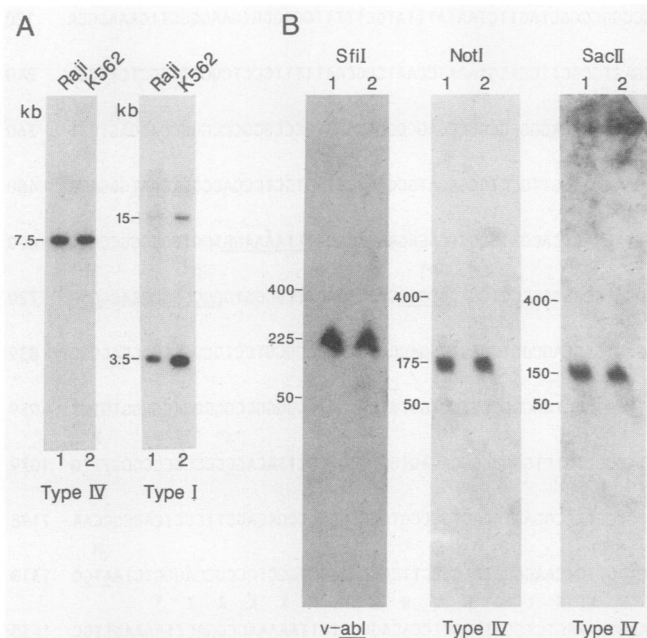


FIG. 3. DNA blot analysis of the human *c-abl* gene. (A) DNA blots containing *Bam*HI digests of Raji DNA (lanes 1) and K562 DNA (lanes 2) after hybridization with the pHab15' exon 1b probe (type IV) or with a 0.46-kb *Eco*RI-*Hinc*II mouse *c-abl* type I cDNA fragment (type I). The latter probe contains sequences derived from exon 1a and the first two common exons of *c-abl*. (B) Hybridization of *abl* probes to large restriction fragments separated by pulsed field electrophoresis. The *Sfi*I blot shows hybridization of probe A28-2-2 (a 0.6-kb *Bam*HI fragment from an intron in the common exon region of *c-abl*) to a 225-kb *Sfi*I fragment. The *Not*I and *Sac*II blots show hybridization of the exon 1b probe to 175-kb *Not*I and 150-kb *Sac*II fragments. Multimers of phage lambda DNA were used as size markers. Lanes: 1, DNA from the pre-B-lymphoid leukemia line RCH-ACV; 2, DNA from the T-cell leukemia cell line SupT3.

common *c-abl* exon (which lies 19 kb downstream of exon 1a [29]) is approximately 200 kb long.

The fact that exon 1b is not amplified in K562 DNA suggests that it lies upstream of the K562 chromosome 9;22 translocation breakpoint (see above). To test whether this was the case, we rehybridized the blots with probe T39-1-2, a 1.0-kb *Eco*RI fragment that maps adjacent to the chromosome 9 breakpoint in K562 cells (16). This probe hybridized to the 225-kb *Sfi*I fragment that was also seen with the common exon probe and to 175-kb *Not*I and 150-kb *Sac*II fragments that had previously hybridized to the exon 1b probe (data not shown; Fig. 4). These results suggest that there is only a single 175-kb *Not*I fragment between exons 1a and 1b, and they confirm that *c-abl* is interrupted by a chromosomal break in K562. Attempts to map the translocation breakpoint in K562 cells more precisely have not yielded conclusive results. In *Sfi*I digests of K562 DNA, a 225-kb fragment is seen with both 5' *bcr* and 3' *c-abl* probes (data not shown). This fragment extends from an *Sfi*I site in the *bcr* region to the site in the 3' part of *c-abl* and comigrates with the germ line 225-kb *Sfi*I *c-abl* gene fragment shown in Fig. 3B. The *Sfi*I site in the *bcr* region maps within 5 kb of the chromosomal break (data not shown), suggesting that the translocation breakpoint in K562 lies closely downstream of exon 1b. In conventional blots we have not found evidence for rearrangements within 8 kb

downstream of exon 1b in K562 DNA, however, and the translocation breakpoint must therefore lie further downstream. A precise localization of the K562 breakpoint awaits the isolation of genomic clones from this region.

The presence of an intron of at least 200 kb in *c-abl* suggests that the gene may be disrupted by the majority of the Philadelphia chromosome translocations in CML. By presenting a large target size, *c-abl* may be especially vulnerable to this type of DNA rearrangement. The proportion of the Philadelphia chromosome translocation breakpoints falling within the large first intron of *c-abl*, however, remains unclear. Three breakpoints have been mapped in the 3' part of the large intron, within approximately 10 kb of the first common exon (17, 23). In the CML cell line K562 the 9;22 breakpoint maps in the 5' part of the large intron (this paper), and we have recently found similar results for the CML line BV173 (C. A. Westbrook, C. M. Rubin, John J. Carrino, Michelle M. Le Beau, A. Bernards, and Janet D. Rowley, submitted for publication). The translocation does not appear to disrupt *c-abl* in all CML DNAs however. Probe T39-2-2, which represents a second CML translocation breakpoint and which was analyzed because it hybridized to DNA fragments that were not amplified in K562 DNA, hybridized to the same 75-kb *Sfi*I fragment as the exon 1b probe, but to different *Not*I and *Sac*II fragments (Fig. 4). The T39-2-2 Philadelphia chromosome breakpoint, therefore, maps upstream of exon 1b. It remains to be seen how often Philadelphia chromosome breakpoints occur upstream of the entire *c-abl* gene and whether in these cases secondary DNA rearrangements are required to form the hybrid *bcr-abl* transcription unit.

The presence of an intron of at least 200 kb in the *c-abl* gene poses several questions. Foremost among these is whether the entire 230-kb gene (if there is only one 175-kb *Not*I fragment between exons 1a and 1b) is contiguously transcribed. This would take over 2 h, given a polymeriza-

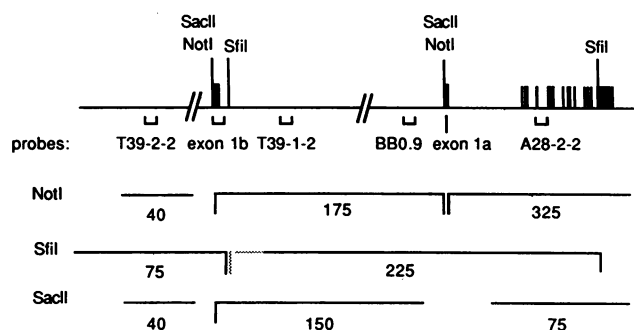


FIG. 4. Map of the human *c-abl* gene showing the 200-kb first intron. Probes used: A28-2-2 (Fig. 3B legend); BB0.9, a 0.9-kb *Bam*HI-*Bgl*II fragment that maps approximately 10 kb upstream of exon 1a; a 1.0-kb *Eco*RI fragment (T39-1-2) that represents the chromosome 9 translocation breakpoint in K562 cells (16); the human *c-abl* exon 1b probe described in the legend of Fig. 1; T39-2-2, a 1.4-kb *Eco*RI fragment representing the chromosome 9 translocation breakpoint from the DNA of a CML patient. The relative order of these probes is shown in the top part of the map. The *Not*I, *Sfi*I, and *Sac*II fragments that are detected by each probe are indicated below. These fragments are not drawn to scale. The map assumes that probes BB0.9, T39-1-2, and the exon 1b probe detect the same 175-kb *Not*I fragment (see the text). The 225-kb *Sfi*I fragment may extend up to the *Sfi*I site 6 kb downstream of exon 1b, but we cannot exclude the presence of additional small *Sfi*I fragments at this location. This uncertainty is indicated by the dashed end of the 225-kb *Sfi*I fragment. Fragment sizes are in kilobase pairs.

tion rate of RNA polymerase II of 30 nucleotides per second (3). Although contiguous transcription seems likely, because bimolecular splicing in eucaryotes has so far only been observed in trypanosomes (6) and chloroplasts (20), it remains to be proven. A related question is how introns of this size are accurately spliced. Splicing over large distances is not restricted to exon 1b, because in K562 and BV173 cells a *bcr* exon splices to the first common exon of *c-abl* over a similar distance. Perhaps the first common *c-abl* exon has the ability to accept donor splices from far upstream. More interestingly, perhaps some of the specificity of the *bcr* region as a partner of *c-abl* in forming the Philadelphia chromosome arises from the ability of the *bcr* exons to splice over long distances to the common *abl* exon. A final question concerns the reason for the existence of this large intron. Possibly the intron consists of selfish DNA without any function. Alternatively, the intron may harbor more N-terminal coding exons of *c-abl*, the expression of which is restricted to specific cell types or developmental stages. It is also possible that the intron contains a complete, unknown gene. Interestingly, Grosveld et al. have observed splicing of *bcr* exons to a sequence in this intron in K562 cells. When this sequence was used to probe RNA blots, 4.5- and 9.5 kb transcripts were detected (16). Further analysis may reveal whether *c-abl*, like the *Gart* gene in *Drosophila* or the gonadotropin-releasing hormone gene in rats (2, 18), harbors a second gene in its introns.

#### ACKNOWLEDGMENTS

We thank Janet Rowley and Gerard Grosveld for helpful discussion and advice and Gerard Grosveld and John Groffen for the A28-2-2, BB0.9, T39-1-2, and T39-2-2 probes. We also thank Nathaniel Landau for the human recombinant library.

This work was supported by Public Health Service program project grant CA 38497 from the National Cancer Institute to D.B., by Public Health Service grant GM 07190 to C.M.R. from the National Institutes of Health, and by grants from the American Cancer Society, Illinois division, and the University of Chicago Research Foundation to C.A.W. A.B. is supported by the European Molecular Biology Organization. C.A.W. is a Stratton-Jaffé Fellow of the American Society of Hematology.

#### LITERATURE CITED

- Abelson, H. T., and L. S. Rabstein. 1970. Lymphosarcoma: virus induced thymic-independent disease in mice. *Cancer Res.* **30**:2213-2222.
- Adelman, J. P., C. T. Bond, J. Douglass, and E. Herbert. 1987. Two mammalian genes transcribed from opposite strands of the same DNA locus. *Science* **235**:1514-1517.
- Alberts, B., D. Bray, J. Lewis, M. Raff, K. Roberts, and J. D. Watson. 1983. *Molecular biology of the cell*, p. 412. Garland Publishing Inc., New York.
- Ben Neriah, Y., A. Bernards, M. Paskind, G. Q. Daley, and D. Baltimore. 1986. Alternative 5'-exons in *c-abl* mRNA. *Cell* **44**:577-586.
- Bernards, A., J. M. Kooter, P. A. M. Michels, R. M. P. Moberts, and P. Borst. 1986. Pulsed field gradient electrophoresis of DNA digested in agarose allows the sizing of the large duplication unit of a surface antigen gene in trypanosomes. *Gene* **42**:313-322.
- Borst, P. 1986. Discontinuous transcription and antigenic variation in trypanosomes. *Annu. Rev. Biochem.* **55**:701-732.
- Carle, C. F., and M. V. Olson. 1984. Separation of chromosomal DNA molecules from yeast by orthogonal-field-alternation gel electrophoresis. *Nucleic Acids Res.* **12**:5647-5664.
- Collins, S. J., and M. T. Groudine. 1983. Rearrangement and amplification of *c-abl* sequences in the human chronic myelogenous leukemia cell line K-562. *Proc. Natl. Acad. Sci. USA* **80**:4813-4817.
- Dierks, P., A. van Ooyen, M. D. Cochran, C. Dobkin, J. Reiser, and C. Weissmann. 1983. Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit beta-globin gene in mouse 3T6 cells. *Cell* **32**:695-706.
- Dynan, W. S., and R. Tijan. 1983. The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell* **35**:79-87.
- Feinberg, A. P., and B. Vogelstein. 1984. A technique for labeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **137**:266-267.
- Gale, R. P., and E. Canaani. 1984. An 8 kilobase *abl* RNA transcript in chronic myelogenous leukemia. *Proc. Natl. Acad. Sci. USA* **81**:5648-5652.
- Goff, S. P., E. Gilboa, O. N. Witte, and D. Baltimore. 1980. Structure of the Abelson murine leukemia virus genome and the homologous cellular gene: studies with cloned viral DNA. *Cell* **22**:777-785.
- Groffen, J., N. Heisterkamp, F. Reynolds, Jr., and J. R. Stephenson. 1983. Homology between phosphotyrosine acceptor sites of human *c-abl* and viral oncogene products. *Nature (London)* **304**:167-169.
- Groffen, J., J. R. Stephenson, N. Heisterkamp, A. de Klein, C. R. Bartram, and G. Grosveld. 1984. Philadelphia chromosomal breakpoints are clustered within a limited area, *bcr*, on chromosome 22. *Cell* **36**:93-99.
- Grosveld, G., T. Verwoerd, T. van Agthoven, A. de Klein, K. L. Ramachandran, N. Heisterkamp, K. Stam, and J. Groffen. 1986. The chronic myelocytic leukemia cell line K562 contains a breakpoint in *bcr* and produces a chimeric *bcr/c-abl* transcript. *Mol. Cell. Biol.* **6**:607-616.
- Heisterkamp, N., J. R. Stephenson, J. Groffen, P. F. Hansen, A. de Klein, C. R. Bartram, and G. Grosveld. 1983. Localization of the *c-abl* oncogene adjacent to a translocation break point in chronic myelocytic leukemia. *Nature (London)* **306**:239-242.
- Henikoff, S., M. A. Keene, J. W. Fachtel, and W. Fristrom. 1986. Gene within a gene: nested *Drosophila* genes encoding unrelated proteins on opposite DNA strands. *Cell* **44**:33-42.
- Hoffman-Falk, H., P. Einat, B.-Z. Shilo, and F. M. Hoffman. 1983. *Drosophila melanogaster* DNA clones homologous to vertebrate oncogenes: evidence for a common ancestor to the *src* and *abl* cellular genes. *Cell* **32**:589-598.
- Koller, B., H. Fromm, E. Galun, and M. Edelman. 1987. Evidence for in vivo trans splicing of pre-mRNAs in Tobacco chloroplasts. *Cell* **48**:111-119.
- Konopka, J. B., S. M. Watanabe, and O. N. Witte. 1984. An alteration of the human *c-abl* protein in K562 cells unmasks associated tyrosine kinase activity. *Cell* **37**:1035-1042.
- Kozak, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**:283-292.
- Leibowitz, D., K. Schaefer-Rego, D. W. Popenoe, J. G. Mears, and A. Banks. 1985. Variable breakpoints on the Philadelphia chromosome in chronic myelogenous leukemia. *Blood* **166**:239-241.
- Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65**:499-560.
- Rowley, J. D. 1973. A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature (London)* **243**:290-293.
- Schwartz, D. C., and C. R. Cantor. 1984. Separation of yeast chromosome-sized DNA molecules by pulsed field gradient gel electrophoresis. *Cell* **37**:67-75.
- Selden, J. R., B. S. Emanuel, E. Wang, L. Cannizzaro, A. Palumbo, J. Erikson, P. C. Nowell, G. Rovera, and C. M. Croce. 1983. Amplified *Cλ* and *c-abl* genes are on the same marker chromosome in K562 leukemia cells. *Proc. Natl. Acad. Sci. USA* **80**:7289-7292.
- Shtivelman, E., B. Lifshitz, R. P. Gale, and E. Canaani. 1985. Fused transcript of *abl* and *bcr* genes in chronic myelogenous

- leukaemia. *Nature (London)* **315**:550-554.
29. Shtivelman, E., B. Lifshitz, R. P. Gale, B. A. Roe, and E. Canaani. 1986. Alternative splicing of RNA transcribed from the human *abl* gene and from the *bcr-abl* fused gene. *Cell* **47**: 277-284.
30. Wang, J. Y. J., and D. Baltimore. 1983. Cellular RNA homologous to the Abelson murine leukemia virus transforming gene: expression and relationship to the viral sequence. *Mol. Cell. Biol.* **3**:773-779.
31. Witte, O. N., A. Dasgupta, and D. Baltimore. 1980. Abelson murine leukemia virus protein is phosphorylated in vitro to form phosphotyrosine. *Nature (London)* **283**:826-831.